基于多源数据和 Transformer 框架的超短期太阳辐照度预测模型

蒋素琼[1]，许培德[2]，王怀远[3]
(1.福建水利电力职业技术学院 电力工程学院，福建 永安 366000；2.福建水利电力职业技术学院 交通工程学院，福建 永安 366000；3.福州大学 电气工程与自动化学院，福建 福州 350108)

摘 要：光伏发电的间歇性与波动性影响电网稳定，高精度预测太阳辐照度能为电力系统调度提供关键支撑。基于气象数据和天空图像的太阳辐照度预测方法已经被广泛研究，但是对于多尺度时序特征和天空图像特征的挖掘还十分有限，也没有充分利用特征间的相关性。文章提出了基于多源数据和 Transformer 框架的超短期太阳辐照度预测模型。首先，利用长短期记忆神经网络-Transformer 学习多元气象序列的长短期时序变化特征；其次，利用卷积视觉 Transformer 对天空图像的全局信息和局部结构进行建模，提取深层次图像特征；最后，构建基于注意力机制的跨模态特征融合模块，利用图像特征表征时序特征，实现跨模态特征融合，由解码器输出预测结果。实验结果表明，与公认的先进方法相比，该方法的预测性能更优，与仅采用气象数据的情况相比，模型的归一化均方根误差平均降低了 7.4%。模型通过挖掘多源数据(气象时序与天空图像)特征，并利用跨模态注意力机制有效挖掘其关联性，显著提升了预测精度。

关键词：太阳辐照度预测；多源数据；天空图像；长短期记忆神经网络；Transformer；CvT
中图分类号：TK519；TM761 **文献标志码**：A **文章编号**：1671-5292(2025)12-1609-10

0 引言

光伏发电凭借资源丰沛、成本优势和零排放特性，成为推动能源清洁化转型的关键路径。光伏发电输出功率主要取决于太阳辐照度(Global Horizontal Irradiance，GHI)，然而，受复杂非线性气象系统特别是云团运动影响[1,2]，超短期 GHI 呈现显著间歇性与波动性，给电力系统带来挑战。

基于数值天气预报的传统物理模型和历史气象数据的统计模型受到算法原理的限制，已经无法实现更高精度的 GHI 预测[3,4]。近些年，天空成像技术的发展为 GHI 预测提供了更加丰富的数据源[5,6]。基于天空图像的 GHI 预测方法主要分为物理方法和深度学习方法。物理方法依赖人工定义的参数(如红蓝比)，鲁棒性和泛化性不足；深度学习方法能自主完成高效特征提取，融合多源数据，实现高维非线性映射[7]。文献[8]利用长短期记忆神经网络 (Long Short-Term Memory，LSTM)进行预测，实验结果表明，LSTM 比机器学习模型 GHI 预测值的 RMSE 平均降低了 15.1%。然而，LSTM 善于捕捉短期时序依赖，却难于处理长期依赖[9]。Transformer 则摒弃循环结构，利用注意力机制实现全序列点对点交互，捕捉数据中的长距离依赖关系[10]。文献[11]在 LSTM 的基础上利用 Informer 对时间序列进行编码，有效提升了模型的预测精度。在基于多源数据的超短期 GHI 预测中，大多数方法都仅采用了单一模型对时序特征进行编码，导致在处理气象时序数据时不能很好地捕捉到时间序列的短期和长期依赖关系[12,13]。卷积神经网络 (Convolutional Neural Network，CNN)凭借其局部感知与权值共享机制，已成为基于图像 GHI 预测模型的主流框架[14]。文献[15]利用残差网络学习连续 2 帧的天空图像，实现了较高精度的光伏出力预测。CNN 虽然在局部信息建模中有优势，但在学习全局信息时效率较低。而基于 Transformer 的视觉模型通过位置编码保留空间位置信息，能更有效地捕获全局特征[16]。此外，多数方法仅针对单一数据源 (如纯时序或纯图像)建模，未能深入挖掘多源数据(天空图像与多元气象序列)间的跨模态关联，制约了预测精度进一步提升[17]。文献[18]设计了基于视觉 Transformer 和门

收稿日期：2025-06-22。
基金项目：福建省自然科学基金(2022J01113)。
通信作者：许培德(1984-)，男，硕士，教授，研究方向为电力新能源储能技术、光伏功率预测技术。E-mail：15859875552@163.com

控循环单元的集成框架，处理多源数据的多模态特征,构建多层感知机融合模块,实现跨模态特征融合。文献[19]则利用 LSTM 结构将图像特征和时序特征进行多源融合,实现高精度预测。

综上所述，只有高效提取多源数据的潜在特征、挖掘多模态特征相关性,才能有效提升预测精度。因此,本文提出了基于多源数据和 Transformer 框架的超短期 GHI 预测模型。首先,构建了时序编码模块,利用 LSTM-Transformer 学习多元气象序列的长短期时序变化趋势;其次,构建了图像编码模块，利用卷积视觉 Transformer(Convolutional Vision Transformer,CvT) 对天空图像的全局信息和局部结构进行建模，捕捉局部特征的同时更好地理解云图中不同区域之间的全局关联;最后,构建了基于注意力机制的跨模态特征模块，融合图像特征和时序特征，实现提前 60 min(60 Minute-ahead,60 MA)的 GHI 预测。实验结果表明，该方法与公认的先进方法 (State of the Art,SOTA)相比具有更好的预测性能,且在整个预测范围内具有很好的鲁棒性。

1 理论方法

1.1 基于 LSTM-Transformer 时序编码模块

LSTM 与 Transformer 分别在短、长时序建模中各具优势。本文据此构建 LSTM-Transformer 融合架构,对历史 GHI 及气象数据进行时间序列编码，通过记忆单元与多头注意力机制提取多尺度时序特征。

1.1.1 长短期记忆神经网络

LSTM 作为递归神经网络的改进架构，专为时序依赖建模设计。其核心创新在于引入记忆单元，由当前输入 x_t 和前一状态 h_{t-1} 共同决定隐藏层状态 h_t。这一机制克服了传统全连接网络因缺乏历史反馈而无法捕捉时序自相关性的缺陷。LSTM 层的原理为

$$\begin{cases} \boldsymbol{i}_t=\sigma(\boldsymbol{W}_{\mathrm{i}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{i}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{i}}) \\ \boldsymbol{f}_t=\sigma(\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{f}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{f}}) \\ \boldsymbol{o}_t=\sigma(\boldsymbol{W}_{\mathrm{o}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{o}}\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{o}}) \\ \tilde{\boldsymbol{c}}_t=\tanh(\boldsymbol{W}_{\mathrm{c}}\boldsymbol{x}_t+\boldsymbol{U}_{\mathrm{c}}\boldsymbol{h}_{t-1}) \\ \boldsymbol{c}_t=\boldsymbol{f}_t\odot\boldsymbol{c}_{t-1}+\boldsymbol{i}_t\odot\tilde{\boldsymbol{c}}_{t-1} \\ \boldsymbol{h}_t=\boldsymbol{o}_t\odot\tanh(c_t) \end{cases} \tag{1}$$

式中:$\boldsymbol{W}_{\mathrm{i}},\boldsymbol{W}_{\mathrm{f}},\boldsymbol{W}_{\mathrm{o}},\boldsymbol{W}_{\mathrm{c}}$ 分别为输入门、遗忘门、输出门和记忆单元的权重矩阵;$\boldsymbol{U}_{\mathrm{i}},\boldsymbol{U}_{\mathrm{f}},\boldsymbol{U}_{\mathrm{o}},\boldsymbol{U}_{\mathrm{c}}$ 为隐藏层的权重矩阵;$\boldsymbol{b}_{\mathrm{i}},\boldsymbol{b}_{\mathrm{f}},\boldsymbol{b}_{\mathrm{o}}$ 为偏置矢量;σ 为 Sigmoid 激活函数;tanh 为双曲正切激活函数。

1.1.2 Transformer 编码器

Transformer 编码器主要由输入表示和多层编码器层组成,其结构如图 1 所示。

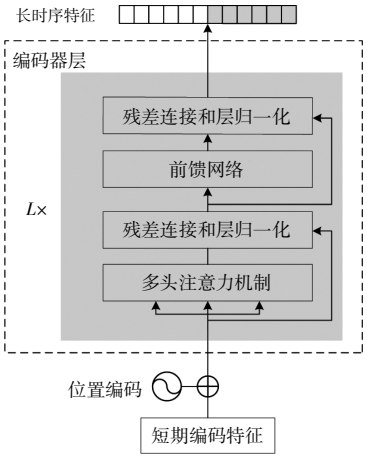


图 1 Transformer 编码器网络结构

Fig.1 Transformer encoder network structure

输入表示包括输入编码和位置编码，其表达式为

$$\boldsymbol{X}=\boldsymbol{X}_{\mathrm{in}}+\boldsymbol{P} \tag{2}$$

$$\begin{cases} \boldsymbol{P}_{(pos,2i)}=\sin\left(\dfrac{\boldsymbol{pos}}{10\,000^{2i/d_{\mathrm{model}}}}\right) \\ \boldsymbol{P}_{(pos,2i+1)}=\cos\left(\dfrac{\boldsymbol{pos}}{10\,000^{2i/d_{\mathrm{model}}}}\right) \end{cases} \tag{3}$$

式中:$\boldsymbol{X}$ 为编码器输入;$\boldsymbol{X}_{\mathrm{in}}$ 为 LSTM 的短期编码特征;$\boldsymbol{P}$ 为位置编码矩阵;$\boldsymbol{pos}$ 为位置索引;d_{model} 为隐藏层维度。

编码器由多层编码器层堆叠而成，编码器层则主要由多头注意力机制、残差连接和前馈网络 3 个部分组成,其表达式为

$$\boldsymbol{X}^{(l)}=\mathrm{EncoderLayer}(\boldsymbol{X}^{(l-1)}),l=1,2,\cdots,L \tag{4}$$

式中:$\boldsymbol{X}^{(l)}$为 l 层编码器层的输出;$\boldsymbol{X}^{(l-1)}$为 $l-1$ 层编码器层的输出;$\boldsymbol{X}^{(0)}=\boldsymbol{X}$ 为初始输入;EncoderLayer() 为编码器层。

Transformer 编码器用于捕捉输入时间序列的上下文关系,从而得到序列的长时序依赖关系。

1.2 基于 CvT 的图像编码模块

CvT 主要由卷积嵌入和卷积 Transformer 编码块两个部分组成,其网络结构如图 2 所示。

由图 2 可见，本文创新性地融合卷积嵌入与

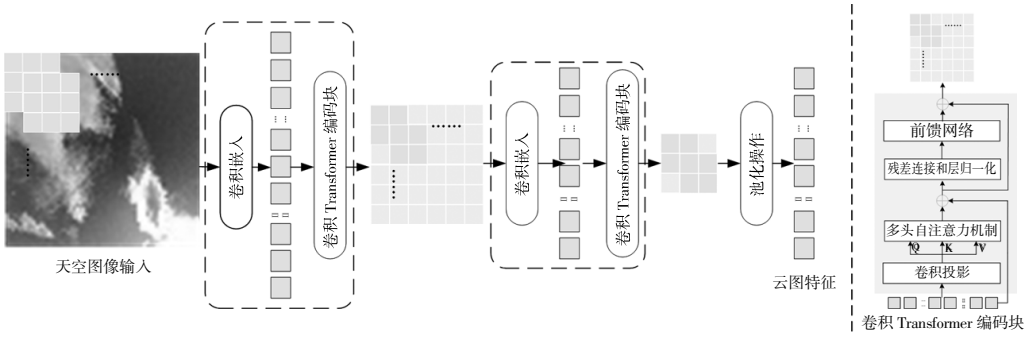


图 2　CvT 网络结构

Fig.2 CvT network structure

投影。通过卷积保留局部信息高效性，利用 Transformer 全局建模优势实现全局特征学习，在不同阶段 CvT 则通过逐步下采样和张量维度提升，使模型能高效提取特征[20]。

1.2.1　卷积嵌入

卷积嵌入是将图像划分为重叠的图像块，通过卷积操作将图像块重塑为二维张量，实现局部特征捕获、序列长度压缩以及维度扩展，完成空间下采样，表达式为

$$\boldsymbol{F}_{\text{conv}}=\text{Conv}(\boldsymbol{O},\boldsymbol{C},S,P) \tag{5}$$

式中：$\boldsymbol{F}_{\text{conv}}$ 为特征图；Conv()为卷积函数；$\boldsymbol{O}$ 为天空图像张量；$\boldsymbol{C}$ 为卷积核；S 为卷积操作滑动步长；P 为填充量大小。

卷积操作的具体过程表达式为

$$\boldsymbol{F}_{\text{conv}}(i,j,m)=\sum_{k=1}^{N}\sum_{l=1}^{N}\boldsymbol{O}(i+k-1,j+l-1)\cdot\boldsymbol{C}(k,l)+\boldsymbol{b} \tag{6}$$

式中：i,j 为输出特征图的位置；m 为输出通道；N 为卷积核大小；$\boldsymbol{C}[k,l]$为卷积核在(k,l)的权重；$\boldsymbol{b}$ 为偏置项。

将嵌入好的特征图展平为序列：

$$\boldsymbol{F}_{\text{flat}}=\text{Flatten}(\boldsymbol{F}_{\text{conv}}) \tag{7}$$

式中：$\boldsymbol{F}_{\text{flat}}$ 为展平序列；Flatten()为展平函数。

1.2.2　卷积 Transformer 编码块

卷积嵌入处理好的天空图像由 CvT 输入到卷积 Transformer 编码块中，过程参见图 2。

首先图像序列经过卷积投影，生成查询、键和值矩阵，然后由多头自注意力机制计算注意力权重，其过程表达式为

$$\begin{cases}\boldsymbol{Q}=\text{ConvProj}(\boldsymbol{F}_{\text{flat}},\boldsymbol{W}_Q)\\ \boldsymbol{K}=\text{ConvProj}(\boldsymbol{F}_{\text{flat}},\boldsymbol{W}_K)\\ \boldsymbol{V}=\text{ConvProj}(\boldsymbol{F}_{\text{flat}},\boldsymbol{W}_V)\end{cases} \tag{8}$$

$$\boldsymbol{A}=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right) \tag{9}$$

$$\text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Concat}(\boldsymbol{head}_1,\boldsymbol{head}_2,\cdots\boldsymbol{head}_h)\boldsymbol{W}_o \tag{10}$$

$$\boldsymbol{F}_{\text{transformer}}=\text{LayerNorm}[\boldsymbol{F}_{\text{flat}}+\text{MultiHead}(\boldsymbol{F}_{\text{flat}})] \tag{11}$$

式中：$\boldsymbol{Q}$ 为查询向量；$\boldsymbol{K}$ 为键向量；$\boldsymbol{V}$ 为值向量；ConProj()为卷积投影；$\boldsymbol{W}_{\text{Q}},\boldsymbol{W}_{\text{K}},\boldsymbol{W}_{\text{V}}$ 为可学习权重矩阵；$\boldsymbol{A}$ 为注意力权重矩阵得分；Softmax()为概率化函数；d_k 为键向量维度；MultiHead()为多头自注意力函数；Concat()为拼接函数；$\boldsymbol{head}_i$ 为第 i 个注意力头；h 为注意力头的数量；LayerNorm ()为层归一化函数。

然后将图像特征序列送入前馈神经网络层，其过程表达式为

$$\text{FFN}(\boldsymbol{F}_{\text{transformer}})=\max(0,\boldsymbol{F}_{\text{transformer}}\boldsymbol{W}_1+\boldsymbol{b}_1)\boldsymbol{W}_2+\boldsymbol{b}_2 \tag{12}$$

$$\boldsymbol{F}_{\text{transformer}}=\text{LayerNorm}[\boldsymbol{F}_{\text{transformer}}+\text{FFN}(\boldsymbol{F}_{\text{transformer}})] \tag{13}$$

式中：FFN()为前馈神经网络函数；$\boldsymbol{W}_1,\boldsymbol{W}_2$ 为权重矩阵；$\boldsymbol{b}_1,\boldsymbol{b}_2$ 为偏置项。

最后通过池化操作得到多源云图特征向量，其过程表达式为

$$\boldsymbol{F}_{\text{final}}=\text{Pool}(\boldsymbol{F}_{\text{transformer}}) \tag{14}$$

式中：$\boldsymbol{F}_{\text{final}}$ 为云图特征向量；Pool()为池化函数。

1.3　基于注意力机制的跨模态特征融合模块

跨模态特征融合整合时序与天空图像特征，虽然两种模块编码方式不同，但均挖掘了影响 GHI 变化的特征。针对现有方法未利用特征相关性的缺陷，本文提出基于注意力机制的跨模态特征融合模块，以图像特征增强时序特征的表征能力，其网络结构如图 3 所示。

时序特征和云图特征分别为 $\boldsymbol{X}_\alpha$ 和 $\boldsymbol{X}_\beta$。为了匹配两种模态特征的张量大小，通过线性映射得到查询向量 $\boldsymbol{Q}_\alpha$、键向量 $\boldsymbol{K}_\beta$、值向量 $\boldsymbol{V}_\beta$。执行跨模

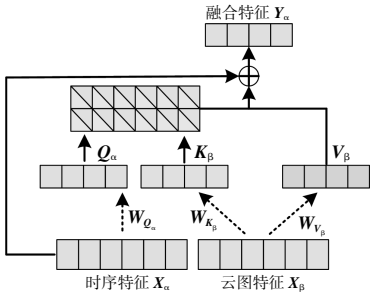


图 3 跨模态特征融合网络结构

Fig.3 Cross-modal feature fusion network structure

态注意力机制的过程为

$$Y_\alpha=\varphi X_\alpha+\psi\mathrm{Softmax}\left(\frac{X_\alpha W_{Q_\alpha}W_{K_\beta}^{\mathrm{T}}X_\beta^{\mathrm{T}}}{\sqrt{d_k}}\right)X_\beta W_{V_\beta},\varphi+\psi=1 \tag{15}$$

式中:W_{Q_α},W_{K_β},W_{V_β}分别为查询向量、键向量和值向量的权重矩阵。

将时序特征作为多头注意力机制中的查询向量，将图像特征作为键向量和值向量，通过计算$Q_\alpha\cdot V_\beta$的得分作为图像特征权重，确定图像特征参与程度。将图像特征权重与值向量相乘,计算得到受到关注的图像特征。再将其与时序特征按照模型参数权重相加得到融合特征。

2 模型整体框架

本文提出的基于多源数据和 Transformer 框架的超短期 GHI 预测模型的整体框架见图 4。

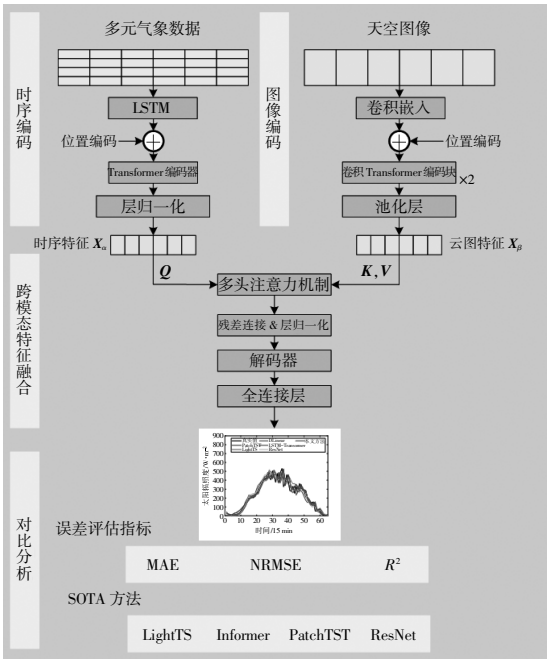


图 4 模型整体框架

Fig.4 The overall framework of the model

首先,通过 LSTM-Transformer 捕获气象序列多尺度时序依赖，采用 CvT 提取天空图像局部-全局特征；其次，利用跨模态注意力融合多源信息;最后,由解码器输出未来 1 h 的 GHI 预测值。同时采用多种误差评估指标，将预测结果与 SOTA 方法进行对比分析。

3 实验设置

以 Inte(R) Xeon(R) Platinum 8350C CPU@ 2.60 GHz,32 GB RAM,NVIDIA GeForce RTX 4090(24 GB)高性能服务器为测试平台,编程语言为 Pytorch 2.1.0 和 Python3.8,深度学习模型全部在 GPU 上进行并行训练。

模型的输入序列长度为 48 步,对未来 1 h 的 GHI 值进行预测,一共预测 6 步,每一步的时间分辨率为 10 min。模型的训练轮次为 12 轮,批量大小为 64,学习率为 0.01,隐藏层维度为 32,损失函数采用均方误差。

3.1 数据描述与预处理

本文采用的数据集源自美国国家可再生能源实验室的太阳辐射研究实验室。实验室安装了 ASI-16 全天成像仪来提供 1 526 像素×1 526 像素 RGB 通道的天空图像，图像的时间分辨率为 10 min。本文采用布朗模型矫正云图畸变。由于原始天空图像具有过高的分辨率，因此在输入模型之前，本文通过下采样的方法将天空图像分辨率调整为 64 像素×64 像素,以此来减少模型的训练时间,部分样本如图 5 所示。同时,实验室还配备了日射强度计、风速计等其他气象传感器来采集 GHI 和气象数据,时间分辨率为 1 min。

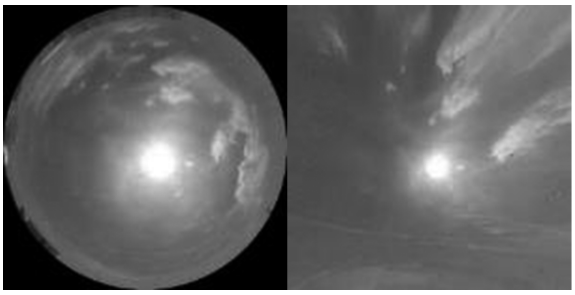


图 5 天空图像预处理

Fig.5 Sky image preprocessing

数据的采集时间为 2019 年 1 月 1 日-2021 年 1 月 1 日,每 10 min 采集一次数据,一共 150 264 个样本。选用其中的 80 000 个样本组成数据集,其中训练集占 70%，验证集占 10%，测试集占

20%。为了避免过拟合,本文只采用训练集进行训练,随后采用验证集实时验证模型的即刻性能,最后单独输入测试集进行测试。

3.2　误差评估指标

采用 3 个广泛使用的评价指标来评估模型的性能:平均绝对误差(Mean Absolute Error,MAE)e_{MA},归一化的均方根误差(Normalized Root Mean Square Error,NRMSE)e_{NRMS},决定系数(R-squared)R^2。

$$e_{MA}=\frac{1}{N}\sum_{i=1}^{N}\left|y_i-\hat{y}_i\right| \tag{16}$$

$$e_{NRMS}=\frac{1}{\max(y)-\min(y)}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2}\times 100\% \tag{17}$$

$$R^2=1-\frac{\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i-\bar{y}_i\right)^2} \tag{18}$$

式中:y_i 为真实值;$\hat{y}_i$ 为预测值;$\bar{y}_i$ 为 y_i 的平均值。

3.3　SOTA 方法

采用不同的 SOTA 模型作为基准模型,与本文方法进行对比,验证本文方法的优越性。

在时间序列预测方面,采用基于 Transformer 的变体模型 Informer 和基于 Transformer 的长时间序列预测模型(Patch Time Series Transformer,PatchTST),以及最近被证明时序建模效果更优的轻量级线性时序模型(Lightweight Time Series Classification with Adaptive Ensemble Distillation,LightTS)。在天空图像处理方面,采用 CNN 架构作为对比模型,并以残差神经网络(Residual Neural Network,ResNet)为核心主干,该模型通过残差学习机制有效克服了深度网络中的梯度消失与梯度爆炸问题。

4　实验与结果分析

4.1　整体性能分析

实验设计采用分阶段对比策略。首先输入多元气象数据,对比 SOTA 时序模型与本文时序编码方法的性能;随后引入主流图像处理模型,在相同跨模态融合框架下,验证天空图像对预测的影响,以证明本文图像编码方法的有效性。各模型预测范围内的误差评估指标平均值如图 6 所示。

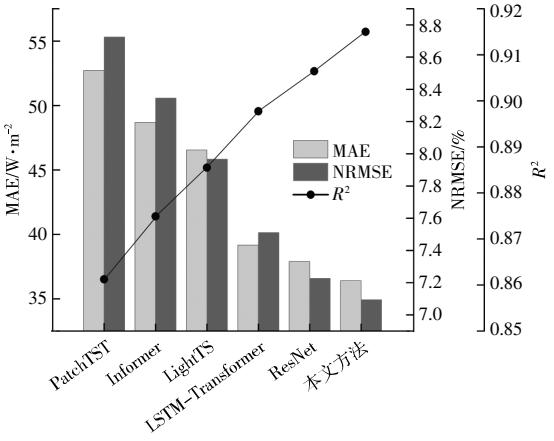


图 6　平均误差评估结果

Fig.6 Average error evaluation results

本文提出的方法与 SOTA 方法的结果对比如表 1,2 所示。

表 1　误差评估结果

Table 1 Error evaluation results

数据集类型	模型	MAE/W·m⁻²						NRMSE/%					
		10 MA	20 MA	30 MA	40 MA	50 MA	60 MA	10 MA	20 MA	30 MA	40 MA	50 MA	60 MA
气象数据	PatchTST	24.32	40.71	51.25	60.18	72.23	81.14	4.94	6.91	8.45	9.57	10.69	12.08
	Informer	31.51	37.98	46.03	50.88	58.24	65.61	5.61	6.92	8.04	9.05	10.11	10.72
	LightTS	24.86	30.76	38.31	42.81	49.06	55.31	5.17	6.46	7.73	8.68	9.69	10.03
	LSTM-Transformer	24.36	31.02	34.33	43.41	47.88	53.21	5.04	6.33	7.31	8.21	9.08	9.44
多源数据	ResNet	23.26	29.38	33.22	41.88	46.25	51.45	4.86	6.21	7.02	7.92	8.83	9.17
	本文方法	21.87	28.32	31.66	40.03	44.27	49.39	4.72	5.79	6.79	7.74	8.34	8.66

表 2　误差评估结果 R^2

Table 2 Error evaluation results R^2

数据集类型	模型	R^2					
		10 MA	20 MA	30 MA	40 MA	50 MA	60 MA
气象数据	PatchTST	0.954	0.915	0.874	0.834	0.793	0.742
	Informer	0.948	0.921	0.879	0.862	0.819	0.787

续表 2

数据集类型	模型	R^2					
		10 MA	20 MA	30 MA	40 MA	50 MA	60 MA
气象数据	LightTS	0.943	0.923	0.891	0.873	0.862	0.829
	LSTM-Transformer	0.957	0.935	0.923	0.885	0.858	0.851
多源数据	ResNet	0.959	0.936	0.914	0.891	0.862	0.848
	本文方法	0.967	0.953	0.934	0.903	0.891	0.869

由表 1,2 可见，对于处理多元气象数据的时序模型,PatchTST,Informer,LightTS 的 NRMSE 在不同时间窗口（10~60 MA）上的误差均高于 LSTM-Transformer。这表明 LSTM-Transformer 在捕捉气象数据的短、长期依赖关系方面更为有效，解决了传统模型在多尺度时序依赖建模上的不足；在输入多源数据的场景下,ResNet 在 10~40 MA 的时间窗口上表现相对较好，但在 50 MA 和 60 MA 时误差明显增加，本文方法在所有时间窗口中均展现出最低的 NRMSE。与 ResNet 相比,在 10~60 MA,NRMSE 分别降低了约 6.72%,6.91%,3.83%,2.27%,5.55%,5.56%。这表明 CvT 能够更有效地提取和利用天空图像全局和局部特征,优于单一的 CNN 架构。因此,评估分析结果证明了本文所提方法的优越性。

除了利用误差评估指标外，本文还利用曲线可视化的方法对模型的预测效果进行分析,GHI 预测结果曲线如图 7 所示。

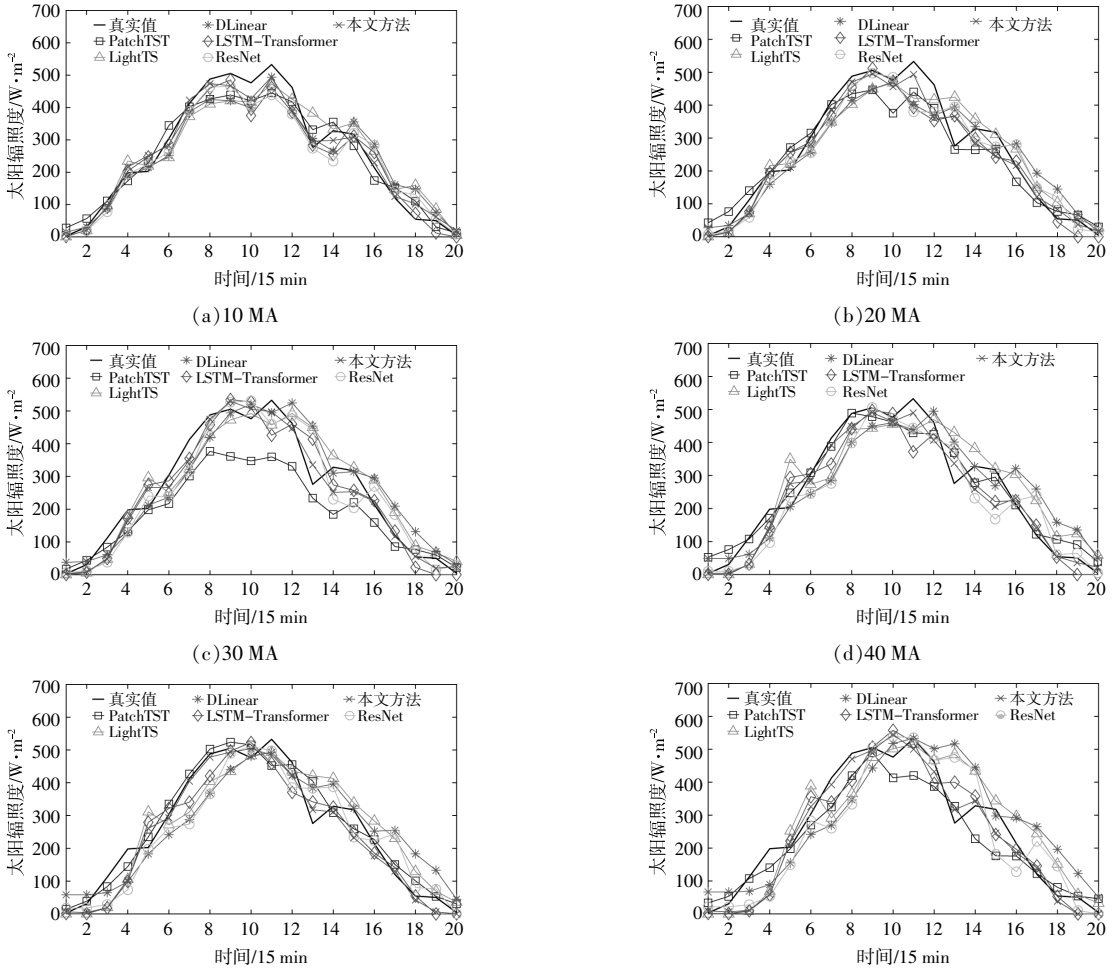
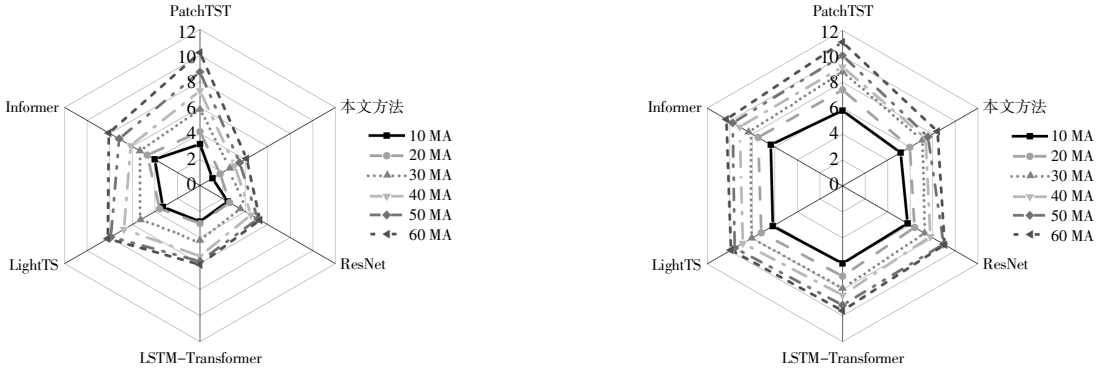


图 7 预测值与真实值对比

Fig.7 The comparison between the predicted value and the real value

由图 7 可见,在 10~30 MA,所有模型预测值的变化趋势与实际值基本上一致，但随着预测范围的扩大,其他模型的预测效果变差,拟合曲线的误差增加；本文所提方法预测曲线可以在整个预测周期内很好地贴合真实值曲线，取得较好的预测效果。

将天气条件分为晴天和阴雨两种情况，验证模型在不同天气条件下的预测效果，并将本文所提方法与 SOTA 模型进行对比分析，结果如图8所示。



(a)晴天　　(b)阴雨

图 8　晴天和阴雨天气下模型预测性能 NRMSE 对比

Fig.8 Comparison of model prediction performance NRMSE in sunny and rainy weather

由图 8 可见，不同预测模型在晴天的预测效果皆优于阴雨天气。当晴天时，所有模型的 NRMSE 都较低,所有方法的误差比较接近,具有较好预测性能。当阴雨天气时,太阳辐射会受到云层和雨水的遮挡,时序模型的预测精度明显下降,不能很好地应对气象突变事件。这就需要天空图像处理模块来提取云层信息,ResNet 是从局部进行图像特征提取,从结果来看有一定效果。本文所提方法则是从局部到全局两个角度挖掘图像信息,比起 ResNet 提取的图像特征更加丰富,更具有竞争力,因此预测误差更小。从整体上看,时序特征主导了 GHI 预测精度,图像特征起到一定的辅助作用,有效提升了 GHI 预测精度。

4.2　跨模态特征融合分析

通过 4.1 节中的分析可知，引入跨模态特征融合可以有效利用多源数据，联合学习时序和图像特征,提升预测精度。为了验证其改进效果,构建了不同类型的主流多模态特征融合策略，替换本文方法中的跨模态特征融合模块进行对比分析，其中包括直接拼接、多层感知机(Multilayer Perceptron,MLP)[7]和 LSTM[19]。直接拼接是将多模态特征向量进行首尾拼接，直接映射得到预测结果。MLP 则是通过 MLP 层将多模态特征融合,由解码器输出预测结果。LSTM 则是将提取后的图像特征和时序特征直接送入 LSTM 层进行预测,实验结果如表 4 所示。

表 4　跨模态特征融合结果评估 NRMSE

Table 4 Evaluation of NRMSE by cross-modal feature fusion results

跨模态特征融合方法	NRMSE/%					
	10 MA	20 MA	30 MA	40 MA	50 MA	60 MA
直接拼接	5.09	6.41	7.28	8.29	9.14	9.51
MLP	4.81	6.18	6.94	7.85	8.69	9.07
LSTM	4.82	6.09	6.88	7.81	8.62	8.93
本文方法	4.72	5.79	6.79	7.74	8.34	8.66

由表 4 可见，直接拼接方法的预测精度明显低于其他融合方法，因为简单的拼接无法挖掘多模态特征的耦合相关性。而 MLP 和 LSTM 的预测性能相较于直接拼接明显提升，其预测效果与本文方法较为接近。这是因为 MLP 和 LSTM 同样充分利用多模态特征的相关性。在已验证的融合方法中，基于注意力机制的跨模态特征融合方法更适用于多源数据的超短期 GHI 预测。

4.3　消融实验

消融实验是验证模型各组成部分贡献的有效手段。本文选取了 LSTM,Transformer,LSTM-Transformer,LSTM-Transformer-CvT 以及本文的方法共 5 种模型进行对比分析。本文方法在不同模块中利用了不同模态的数据,LSTM,Transformer 和 LSTM-Transformer 仅使用多元气象数据,而 LSTM-Transformer-CvT 及其他参照方

法则同时采用气象数据与天空图像数据。为进一步验证跨模态特征融合模块的有效性,LSTM-Transformer-CvT 采用了直接拼接的直接特征融合策略,以对比本文所提出的融合机制,实验结果如表 5 所示。

表 5 消融实验结果评估 NRMSE

Table 5 Evaluation of the ablation experiment results using NRMSE

模型	NRMSE/%					
	10 MA	20 MA	30 MA	40 MA	50 MA	60 MA
LSTM	4.94	6.92	8.13	8.55	9.34	9.91
Transformer	5.14	6.51	7.63	8.39	9.25	9.74
LSTM-Transformer	5.04	6.33	7.31	8.21	9.08	9.44
LSTM-Transformer-CvT	5.09	6.41	7.28	8.29	9.14	9.51
本文方法	4.72	5.79	6.79	7.74	8.34	8.66

由表 5 可见,LSTM 在 10 MA 预测中表现出色,印证了其在捕捉短期时序依赖方面的有效性。因此,融合了 LSTM 和 Transformer 的组合模型拥有长短期时序建模的能力,其预测性能显著优于单一的 LSTM 和 Transformer。然而,在引入 CvT 模块后,模型的预测性能出现了明显下降,这表明若仅添加 CvT 及图像数据而未构建有效的多模态融合机制,反而会导致特征冗余,从而干扰模型预测。由此可见,在引入 CvT 的同时必须辅以合适的跨模态融合策略。本文的方法在此基础上实现了有效融合,相比 LSTM-Transformer 模型得到了进一步提升。因此,本文方法的每个部分都有助于提升 GHI 预测性能。

4.4 敏感性分析

4.4.1 输入序列长度

较长的输入序列能够为 GHI 预测提供更丰富的可学习特征。然而,过长的输入序列也可能与当前的 GHI 预测无关。因此,过短的输入序列会导致特征不足,过长则特征冗余。由于输入序列长度对模型的预测性能十分重要,因此选取4 种输入序列长度进行实验,然后采用网格法在数据集上进行实验验证,结果如图 9 所示。

由图 9 可见,当输入序列长度为 48 时,本文方法在所有预测点都取得了更为优秀的结果。此外,无论采用何种输入序列长度,模型的预测性能都很稳定,表明本文方法在常用的输入序列长度区间内保持了良好的鲁棒性。

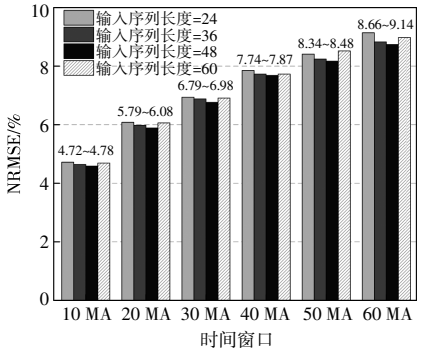


图 9 输入序列长度对预测性能的影响

Fig.9 The influence of input sequence length on prediction performance

4.4.2 图像分辨率

图像分辨率对特征学习的影响与模型任务及特征作用密切相关。尽管低分辨率图像可能损失部分细节特征,但可显著提升训练效率。鉴于本文将图像特征作为辅助输入,过高的分辨率可能因特征冗余导致预测性能下降。实验设置 4 组分辨率(64×64,80×80,100×100,128×128),在相同条件下进行 10 次重复实验取均值,结果如图 10 所示。

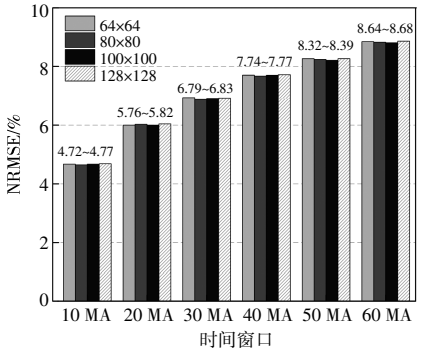


图 10 图像分辨率对预测性能的影响

Fig.10 The influence of image resolution on prediction performance

由图 10 可见,因天空图像特征在本文方法中作为辅助信息,分辨率变化对预测结果影响有限(NRMSE 波动<0.1%),表明模型对分辨率差异具有强鲁棒性;同时,分辨率提升将显著增加训练耗时。综合效率与性能,将原始图像下采样至 64×64 进行 GHI 预测。

5 结论

随着光伏发电对电网渗透率的持续提升,系统稳定性面临严峻挑战,高精度 GHI 预测对其安全运行至关重要。云层分布与运动是影响辐照度

的直接物理因素，而气象数据中隐含的时序特征与天空图像特征存在显著映射关联。本文据此提出基于多源数据和 Transformer 框架的超短期 GHI 预测模型，融合多源数据的跨模态特征实现 GHI 预测,提升预测精度。实验表明,在 10~60 MA 预测时间内，本文模型在 MAE,NRMSE 和 R^2 指标上均显著优于 SOTA 方法。根据误差评估结果可以得出以下结论。

①基于 LSTM-Transformer 的时间编码模块有效地提高了多尺度时序特征提取能力，同时与其他时间序列模型进行比较，具有更高的精度和鲁棒性。

②为了对天空图像进行全局信息和局部结构建模，构建了基于 CvT 的天空图像编码模块,结果证明这种兼顾全局信息和局部结构的特征提取方式比仅关注单一方面的模型（ResNet）更有优势。

③为了能够兼顾图像特征和时间序列特征,本文提出了基于注意力机制的跨模态特征融合方法,融合时间序列特征和天空图像特征,很好地利用了它们之间相关性去提升模型预测精度。

未来可以纳入更多的多源多模态数据，包括卫星云图以及周边光伏电站的空间数据等，为模型提供更多的可学习特征。同时加强图像处理和云动力学的研究,学习云运动机理,引入云动力学模型,从而更准确地挖掘云层的变化特征,提升预测精度。

参考文献:

[1] 吴家葆,曾国辉,张振华,等.基于 K-means 分层聚类的 TCN-GRU 和 LSTM 动态组合光伏短期功率预测[J].可再生能源,2023,41(8):1015-1022.

[2] 李豪,马刚,李天宇,等.基于时空相关性的短期光伏出力预测混合模型 [J]. 电力系统及其自动化学报,2024,36(5):121-129.

[3] Ghimire S,Deo R C,Casillas-Pérez D,et al.Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms [J].Applied Energy,2022,316:119063.

[4] 邱明,鲁冠军,吴昊天,等.基于数据清洗与组合学习的光伏发电功率预测方法研究[J].可再生能源,2020,38(12):1583-1589.

[5] Liu J,Zang H,Cheng L,et al.A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting[J].Applied Energy,2023,342:121160.

[6] 王迎春,王昱栋,刘洋,等.基于卫星遥感多光谱云图的生成式海上超短期光伏功率预测[J].控制与决策,2025,40(4):1136-1144.

[7] 路志英,周庆霞,李鑫,等.基于地基云图图像特征的光伏功率预测[J].电力系统及其自动化学报,2020,32(8):70-76.

[8] Zhang J,Yan J,Infield D,et al.Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model[J].Applied Energy,2019,241:229-244.

[9] Limouni T,Yaagoubi R,Bouziane K,et al.Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model [J].Renewable Energy,2023,205:1010-1024.

[10] 盛瑞祥,张啸宇.基于概率 TCN-Transformer 的短期光伏功率预测模型[J].综合智慧能源,2024,46(11):10-18.

[11] Cao Y,Liu G,Luo D,et al.Multi-timescale photovoltaic power forecasting using an improved Stacking ensemble algorithm based LSTM-Informer model [J].Energy,2023,283:128669.

[12] Xu S,Zhang R,Ma H,et al.On vision transformer for ultra-short-term forecasting of photovoltaic generation using sky images[J].Solar Energy,2024,267:112203.

[13] Cheng L,Zang H,Wei Z,et al.Short-term solar power prediction learning directly from satellite images with regions of interest [J].IEEE Transactions on Sustainable Energy,2021,13(1):629-639.

[14] Wen H,Du Y,Chen X,et al.Deep learning based multistep solar forecasting for PV ramp-rate control using sky images [J].IEEE Transactions on Industrial Informatics,2020,17(2):1397-1406.

[15] Zhang R,Ma H,Saha T K,et al.Photovoltaic nowcasting with bi-level spatio-temporal analysis incorporating sky images [J].IEEE Transactions on Sustainable Energy,2021,12(3):1766-1776.

[16] Yuan L,Chen Y,Wang T,et al.Tokens-to-token vit: Training vision transformers from scratch on imagenet [A].2021 IEEE/CVF International Conference on Computer Vision (ICCV)[C].Montreal:IEEE,2022.538-547.

[17] Jiao X,Li X,Lin D,et al.A graph neural network based

deep learning predictor for spatio-temporal group solar irradiance forecasting[J].IEEE Transactions on Industrial Informatics,2021,18(9):6142-6149.

[18] Xu S,Zhang R,Ma H,et al.On vision transformer for ultra-short-term forecasting of photovoltaic generation using sky images[J].Solar Energy,2024,267:112203.

[19] Yao T,Wang J,Wu H,et al.Intra-hour photovoltaic generation forecasting based on multi-source data and deep learning methods [J].IEEE Transactions on Sustainable Energy,2021,13(1):607-618.

[20] Wu H,Xiao L,Yue C,et al.Cvt:Introducing convolutions to vision transformers [A].2021 IEEE/CVF International Conference on Computer Vision (ICCV)[C].Montreal:IEEE,2022.22-31.

Ultra-short-term solar irradiance prediction model based on multi-source data and Transformer framework

Jiang Suqiong[1], Xu Peide[2], Wang Huaiyuan[3]

(1.School of Electrical Engineering, Fujian College of Water Conservancy and Electric Power, Yong'an 366000, China; 2.School of Transportation Engineering, Fujian College of Water Conservancy and Electric Power, Yong'an 366000, China; 3.School of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: The intermittence and volatility of photovoltaic power generation threaten the stability of the power grid, and high-precision prediction can provide key support for power system scheduling. Solar irradiance prediction methods based on meteorological data and sky images have been widely studied, but the mining of multi-scale temporal features and sky image features is still very limited, and the correlation between features is not fully utilized. Therefore, this paper proposes an ultra-short-term solar irradiance prediction model based on multi-source data and Transformer framework. First, the long-term and short-term memory neural network-Transformer is used to learn the long-term and short-term temporal variation characteristics of multivariate meteorological sequences. Secondly, the convolution visual Transformer is used to model the global information and local structure of the sky image, and the deep image features are extracted. Finally, a cross-modal feature fusion module based on attention mechanism is proposed, which uses image features to represent temporal features, realizes cross-modal feature fusion, and outputs prediction results by decoder. Experimental results show that the proposed method has better prediction performance than the state-of-the-art methods. Compared with the case of only using meteorological data, the NRMSE of the model is reduced by 7.4% on average. The model significantly improves the prediction accuracy by mining the features of multi-source data (meteorological time series and sky images) and using the cross-modal attention mechanism to effectively mine their correlation.

Keywords: solar irradiance prediction; multi-source data; sky image; long short-term memory neural network; Transformer; CvT